STREAMLINING OF THE PROCESS OF DECISION-MAKING IN REAL ESTATE MANAGEMENT WITH INCOMPLETE INFORMATION

Małgorzata Renigier-Biłozor, PhD
Faculty of Geodesy, Geospatial and Civil Engineering
University of Warmia and Mazury in Olsztyn
e-mail: malgorzata.renigier@uwm.edu.pl

Andrzej Biłozor, PhD
Faculty of Geodesy, Geospatial and Civil Engineering
University of Warmia and Mazury in Olsztyn
e-mail: abilozor@uwm.edu.pl

Agnieszka Napiórkowska-Baryła, PhD
Faculty of Econonomics
University of Warmia and Mazury in Olsztyn
e-mail: agnieszka.baryla@uwm.edu.pl

Abstract

This paper presents a streamlined sub-system of decision-making in a real estate market with incomplete data. As we currently observe, various entities collect data and use databases, which entails a problem with their quality and completeness. This results from the specifics of the real estate market, particularly from the nature of the available information, access to it and integral uncertainty.

In the first part of this paper, we will present substantive guidelines for the development of a procedure for supplementing missing information. Afterwards, in order to verify the feasibility and effectiveness of the procedure, an implementation simulation will be conducted on the selected example. We would like to emphasize that all decisions are made under the conditions of an information gap.

1. Introduction

In view of the growing significance of real estate and real estate markets in the social, economic and business sphere, access to reliable, explicit and complete information is very important and often constitutes a precondition for achieving a set goal. What makes the problem more difficult to solve is that there are still no unified systems of information about real estate, while the collection and processing of own databases is becoming a universal phenomenon, both among individual and institutional market players. There are usually information gaps in databases, which causes problems connected with grouping and choosing similar properties, the selection of representative objects,

valuation of real estate, etc. These gaps results from many reasons, including: human error, difficult access to data, and uncertainty prevailing in the real estate market. Moreover, information gaps observed in the real estate market create an absorptive space for information conformity (BRZEZICKA, WISNIEWSKI 2014), and this phenomenon has an impact on the level of prices in the real estate market and the decisions of market players.

The lack of reliable information translates into limited and incomplete knowledge about real estate and, as a consequence, into problems in making optimal decisions (RENIGIER et al. 2014). One of the methods which takes into account the specifics of information relating to the real estate market is a method based on the rough set theory. This theory is used for the examination of impreciseness, generality and uncertainty in the process of data analysis, commonly occurring in the real estate market. In order to increase flexibility in data exploration, the authors propose the integration of the assumptions of the classical rough set theory with the valued tolerance relation (based on the assumptions of fuzzy logic) to make it possible to analyze data expressed in various forms.

2. Decision-making under the conditions of uncertainty

A need to make a decision or to choose a course of action appears when a human faces a problem. Therefore, making a decision means the process of selecting a particular course of action as a way of solving a particular problem. The basic assumption behind decision-making is the rationality of a decision-maker, underpinning the so-called rational and normative concepts of decision-making results, which do not take into account behavioral factors (BOLESTA–KUKUŁKA 2003). To streamline the decision-making process as much as possible, models being simple constructs may be used, which present a simplified image of the examined fragment of reality based on the elimination of one of the elements (features, relations) insignificant for the given goal or at the given stage of examination (PENC 2001). BOLESTA–KUKUŁKA (2003) distinguishes three basic models of rational decision-making processes:

1) model of the mono-criterial decision-making process, which involves the determination of one criterion for evaluating decision variants,
2) multi-criterial model, which involves the identification of criteria for evaluating decision variants and assigning weights to them,
3) two-step valuation model, in which it is proposed to evaluate the generated variants not only as regards their effectiveness, but also as regards feasibility and rationality.

Regardless of the adopted model, a decision-making process, defined as a well-thought-through (rational) sequence of actions, starting from the recognition of a decision problem and setting of goals, proceeds according to Fig. 1).
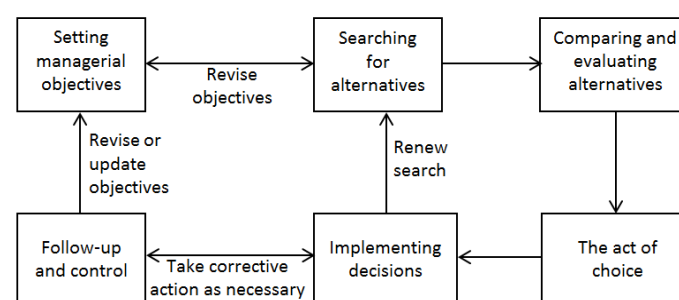
Fig. 1. Decision-making process. *Source*: HARRISON (1996).

However, the rationality of selections is limited, due to such factors as the imperfection and incompleteness of information describing problems, complexity of problems, possibility to process information, time for making a decision, or conflicts between selections and objectives of the organization (HATCH 2002). Depending upon how well a decision-maker is equipped with information and, as a consequence, to what extent he/she knows a decision-making situation, we distinguish three main classes of decision-making situations (BĄK, NALEPKA 2006):

– decision-making under the conditions of certainty (deterministic) – when the situation and causal dependencies are known,

– decision-making under conditions of risk (probabilistic) – when at least one of the circumstances is unknown, but the likelihood of its or their occurrence is known and measured,
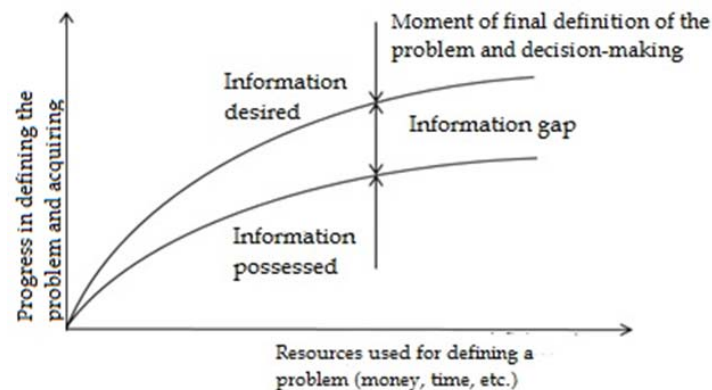– decision-making under conditions of uncertainty (strategic) – when a decision-maker does not know all the circumstances, and it is not possible to assess the likelihood of their occurrence.

Karwacki and Konarzewska (1997) and Kofler (1993) additionally identify:
– decision-making under conditions of incomplete (partial) information.

The quantity of information undoubtedly has an impact on the decision-making process. The difference between the information desired by a decision-maker and those he/she can acquire by incurring the various outlays towards the realization of this goal is called an "information gap" (Fig. 2). It is defined as a sense of the lack of information, knowledge or understanding of the problem, which are necessary (or at the very least useful) for making a decision.

Fig. 2. Concept of an "information gap" in the decision-making process. *Source*: Bąk and Nalepka (2006), after Rabin and Jackowski (1988).

A prevailing number of decisions made are strategic decisions made under conditions of uncertainty. This is also the most important element of activity of investors and managers, which becomes especially significant in the case of high-value tangible investments (Konowalczuk, Ramian 2013), including investments financed from the budget of the State Treasury or local government units (Radzewicz, Walacik 2014; Walacik 2014). A decision is made resulting from the analysis, both of the organization (investor) and the environment, with the aim of specifying the extent of a strategic gap, being a derivative of an information gap, and to minimize it (Fig. 3). The higher the uncertainty level, the more difficult it is to make a decision.
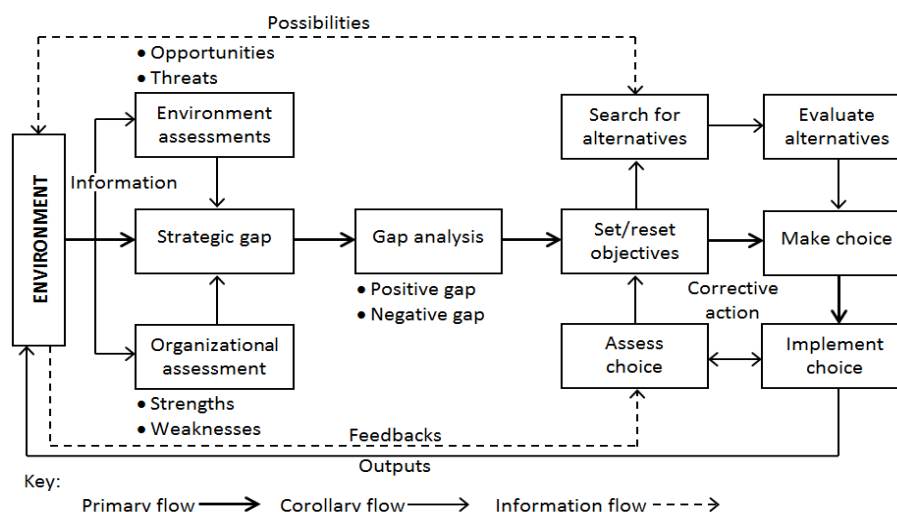
Fig. 3. Strategic decision-making process. *Source*: Harrison (1996).

Decisions in real estate management are, for the most part, made under uncertain situations. Hence, the most important is that the whole process of decision-making should be assumed in such a way so as to reduce the degree of uncertainty and the risk of potential losses if it is not possible to fully eliminate the risk. This can be achieved, among others, by increasing access to consistent and reliable information (RENIGIER-BIŁOZOR 2013). An information gap in real estate databases may be filled by acquiring additional data and information through the simulation of the spatial distribution of variables, in such works as CELLMER, SZCZEPANKOWSKA (2014), CELLMER et al. (2014).

There are no comprehensive and effective IT systems designed for real estate management and analysis relating to real estate. Management of 21st century real estate is not possible without effectively and efficiently functioning systems of collecting and processing information. Real estate management is a continuous process of making decisions, for which it is indispensable to use properly prepared and edited information.

3. Model of decision-making in the real estate market with a shortage of data

In this article, a simplified decision-making model of supplying missing data has been prepared. The presented procedure has been elaborated upon in the modular structure (Scheme 1). Such an approach guarantees methodological openness. The procedure may be implemented including all the modules, or omitting some of them depending on the purpose for which the database is used. The methodology was based on the assumptions of the rough set theory and constitutes a development of the solutions presented in the work by RENIGIER-BIŁOZOR (2010).

In each phase of the procedure, the specific tasks to be performed in subsequent modules have been set. Module I concerns the purpose of utilizing a database. At this stage, the utilization purpose is defined, e.g. investment consultancy or selection of representative properties, etc., as well as the scope of actions required. Module II concerns the creation of a database according to the work order. At this stage, it is necessary to identify the places where there is no data, or the piece of data is not very reliable. Text variables are then coded in the quantity form. Module III contains a decision-making table. Here it is necessary to classify the attributes into conditional and decisional according to the Bool's assumption: if condition… then decision. The domains are then assigned to individual attributes.

Moving on to module IV, one must group the acquired data. This procedure assumes the grouping of individual values of attributes according to the non-differentiability relation. To this end, in accordance with the rough set theory, it is necessary to use a formula of the valued tolerance relation, which enables one to set top and bottom approximation of a set with a varying degree of the non-differentiability relation. The mathematic formula of this relation is:

$$R_j(x,y) = \frac{\max\left(0, \min\left(c_j(x), c_j(y)\right) + k - \max\left(c_j(x), c_j(y)\right)\right)}{k} \qquad (1)$$

where:

$R_j(x, y)$ — the relation between two sets with a membership function [0,1],

$c_j(x), c_j(y)$ — the variable of the analyzed property,

k — the coefficient adopted for a given property attribute.

A sensitive point at this stage is setting the k coefficient, which presents the range of variability of the set. In the procedure concerned, it is permissible to set the k coefficient on the basis of the database being created, or to adopt this coefficient *a priori* when, for example, we have a larger database at our disposal.

The next module (V) concerns the determination of the similarity (non-differentiability) of the analyzed objects against the k coefficient, separately for each attribute. Considering Formula 1, we have to remove negative results assuming their maximum differentiability. Next, we have to determine the relative similarity of all attributes in the set by calculating matrices of sums of similarity (non-differentiability) using the valued tolerance relation according to the formula:

$$R_j(x,p) = \max\left(\sum_{j=1}^{n} R_j(x,p)\right) \qquad (2)$$

where: R_j is the valued tolerance relation, x is the analyzed property's attribute, p is the attribute in the conditional segment of the investigated decision rule, and n is the number of property attributes in the conditional segment of the decision rule.

Module I

Start — defining the purpose of data base utilization

Module II — identifying missing data — Creation of database — coding text variables

Module III — assigning domains to attributes — Preparation of decision-making table — calculating k coefficient on the basis of the provided database.

Module IV — Grouping of attributes — calculating coefficients of data variation.

optionally – calculating universal k coefficient - *a priori*.

Module V — deleting negative results as totally differentiable — calculating similarity (non-differentiability) matrix against k coefficient — Determination of similarity (non-differentiability) — calculating a matrix of sums of similarity (non-differentiability) using the valued tolerance relation.

deleting data with gaps in records.

Module VI — calculating abstraction classes for the given similarity. — Setting decision-making rules — determining a minimum level of similarity

Module VII — elimination of redundancies of similar objects. — Determination and edition of missing attributes — disclosing the lack of similar objects.

Module VIII — Calculation and analysis of the accuracy of approximation

Module IX — Supplying missing data — Decision-making

Scheme 1. Model of decision-making in the real estate market under a shortage of information. *Source*: Own work.

In module VI, decision-making rules are set according to the rough set theory based on Bolo's assumptions (If....then...). At this stage of the procedure, abstraction classes for the given non-differentiability relations are calculated from matrices of conditional and decisional sums. In the next phase, one must set a minimum level of the assumed similarity.

Module VII includes the identification and edition of missing attributes on the basis of the previously calculated quality of approximation. In this phase, feedback to the procedure may appear. This takes place when disclosing of redundancies of objects similar to the objects with missing data, when it is not possible to calculate an clear result. We then have to return to module VI and eliminate any excessive objects (i.e. having the largest data sets) on the basis of the strictest decision-making rules. A similar effect is obtained when disclosing a lack of similar objects with the assumed minimum level of approximate similarity; the minimum level of similarity in module VI should then be reduced.

Module VIII assumes a calculation and analysis of the accuracy of approximation. Generally, according to the assumptions of the rough set theory, it is possible to check the accuracy by calculating the approximation of sets as follows:

if $S_c = \langle U, \widetilde{C} \rangle$ is an approximation space, and any set $C \subseteq U$ then:

$\widetilde{C}$ - bottom approximation X in S_c is a set: $\underline{\widetilde{C}}X = \{x \in U : [x]_{\widetilde{C}} \subseteq X\}$ (3)

while

$\widetilde{C}$ - top approximation X in S_c is a set: $\overline{\widetilde{C}}X = \{x \in U : [x]_{\widetilde{C}} \cap X \neq 0\}$ (4)

This results from the calculation that for any $X \subseteq U$, X is $\widetilde{C}$ - crisp if and only if $\underline{\widetilde{C}}X = \overline{\widetilde{C}}X$, while X is rough, when $\underline{\widetilde{C}}X \neq \overline{\widetilde{C}}X$.

In the case of the approximation of rough sets, the approximation of families of sets may be characterized by numbers using certain coefficients. In the case of families of sets, we can use the following measure:

– accuracy of the approximation of F family in S approximation space, against a set of attributes C:

$$\beta_{\widetilde{C}}(F) = \frac{card\big(POS_{\widetilde{C}}(F)\big)}{\sum\limits_{X_i \in F} card\big(\overline{\widetilde{C}}X_i\big)}$$ (5)

For the above coefficients, the following patterns occur:

– $0 \leq \alpha_{\widetilde{C}}(F) \leq \gamma_{\widetilde{C}}(F) \leq 1$

– if all elements of the F family are C – crisp sets, then: $\alpha_{\widetilde{C}}(F) = \gamma_{\widetilde{C}}(F) = 1$

Module IX ends the decision-making procedure. One must supply missing data (fill in the gaps in the records) with the results from the analyses performed on the basis of the highest approximation, and then use them for making a proper decision in the real estate market.

3.1. Simulation of the implementation of the decision-making model of supplementing incomplete information

Afterwards, the methodological assumptions presented in the first phase will be tested on the exemplary database. The above-mentioned simulation is presented below, with the module structure of the decision model retained.

Module I – the analyzed database covers 120 transactions with residential premises in Olsztyn from 2013, described by 9 attributes (Table 1). The work order concerns supplementing the database with the aim of using it for investment consultancy in the real estate market.

Module II – moving on to the second module, the objects were identified with the missing data (Table 2).

The text variables were then coded and classified according to the following assumptions:

– location: quantitative classification – ordinal scale,
– usable area of real estate: quantitative classification - ratio scale,

---

–   number of rooms: quantitative classification - ratio scale,
–   location on floor: quantitative classification - ordinal scale,
–   height of the building: quantitative classification - ordinal scale,
–   age of the building: quantitative classification – interval scale,
–   appurtenant cellar compartment: quantitative classification – nominal scale,
–   price of real estate: quantitative classification – ratio scale.

Table 1

Exemplary database from the real estate market in Olsztyn

No. trans.	Price	Location	District	Usable area	Number of rooms	Location on floor	Height of building	Age of building	Price
1	135000	Liliowa	Redykajny	29.50	1	4	4	2001	1
2	235000	Grunwaldzka	Grunwaldzkie	75.20	4	0	3	1907	1
3	210000	Hallera	Jaroty	47.20	2	1	4	2000	1
4	220000	Małłków	Śródmieście	58.68	4	0	4	1962	0
5	245000	Profesorska	Podgrodzie	58.40	4	0	3	1956	1
6	205000	Popiełuszki	Generałów	36.89	2	0	4	2011	1
7	170000	Sybiraków	Podleśna	46.90	4	9	11	1970	1
8	176200	Boenigka	Jaroty	48.50	3	4	5	1984	0
9	199620	Popiełuszki	Generałów	44.01	3	1	4	2011	1
10	185000			47.00	3	0	3	1981	1
11	259000	Mroza	Jaroty	72.30	5	0	5	1991	0
12	151196	Puszkina	Zatorze	42.70	3	2	5	1968	0
13	210960	Piotrowskiego	Jaroty	87.90	4	1	5	2010	1
14	205000	Srebrna	Jaroty	45.53	2	4	7	2006	1
15	156250	Dworcowa	Kormoran		3	0	4	1953	0
16	120000	Wyszyńskiego	Kormoran	32.30	3	3	5	1975	0
17	230000	Dolna	Jaroty	45.15	2	1	3	2011	0
18	142000	Zamenhofa	Kętrzyńskiego	34.40	3	0	4	1963	0
19	146286	Popiełuszki	Generałów	33.63	2	2	4	2011	1
20	195000	Dworcowa	Kormoran	58.00	4	2	12	1978	0
21	217000	Orłowicza	Nagórki	59.60	4	5	10	1980	1
22	164975	Dworcowa	Kormoran	48.00	4	10	12	1977	1
23	256000	Sucharskiego	Jaroty	68.60	2	2	5	2005	1
24	208865	Murzynowskiego	Nagórki	60.40	4	3	12	1982	1
25	152000	Zamenhofa	Kętrzyńskiego	34.40	3	2	4	1963	0
….									
120	148243	Gdańska	Podleśna	42.50	3	1	2	1930	0

Source: own study.

Table 2

Table with missing records in the description of the real estate

Trans. No.	Missing attribute
10	location
15	usable area
47	number of rooms
82	price
85	location on floor
116	age of building and appurtenant cellar compartment

Source: own study.

Module III – in this module, data was divided into conditional and decisional attributes. According to the accepted work order, the data will be used for purposes of investment consultancy, so the price (d) will be the decisional attribute (Table 3), while the other attributes are conditional (c_1, c_2, c_3, c_4, c_5, c_6, c_7).

Afterwards, attribute domains were defined according to the following criteria:
– location: 1 – good; 2 – average; 3 – unfavorable,
– usable area of real estate: in square meters,
– number of rooms: nominal value,
– location on floor: 1 – 1st and 2nd, 2 – 3rd and 4th floor, 3 – ground floor and above 4th floor,
– height of building: 1 – up to 4 stories, 2 – up to 6 stories, 3 – over 6 stories,
– age of building: 1 – up to 5 years, 2 – 6 to 10 years, 3 – 11 to 20 years, 4 – 21 to 30 years, 5 – 31 to 50 years, 6 – 51 to 80 years, 7 – above 80 years,
– appurtenant cellar compartment: 1 – yes, 0 – no,
– price of real estate in PLN/m².

Table 3

List of analyzed attributes

Conditional attributes							Decision attribute
c_1	c_2	c_3	c_4	c_5	c_6	c_7	d
Location	Usable area	Number of rooms	Location on floor	Height of building	Age of building	Appurtenant cellar compartment	Price

Source: own study.

Module IV – moving on to module IV, one has to group the available data. In order to use the formula of the valued tolerance relation (Formula 1), which allows one to define the top and bottom approximation of a set with various degrees of the non-differentiability relation, the *k* coefficient was calculated on the basis of the standard deviation of the individual values of attributes (Table 4).

Table 4

Coefficient *k* for various attributes

	Attributes							
	c_1	c_2	c_3	c_4	c_5	c_6	c_7	d
Coefficient *k*	0.68	17.40	0.96	0.90	0.80	0.71	0.50	787.68

Source: own study.

Module V – in module V, the similarity (non-differentiability) of the analyzed data was defined. In the first phase, a similarity (non-differentiability) matrix was calculated against the *k* coefficient set above, separately for each attribute (exemplary matrix for the space – Table 5).

Table 5

Matrix of similarities for the usable area attribute

Sum	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	120
1	7.00	2.00	3.00	1.00	2.00	2.57	2.00	1.00	2.15	2.00	0.00	0.22	1.00	2.06	1.00	2.24
2	2.00	7.00	2.00	3.03	5.01	3.00	3.00	0.00	2.00	3.00	1.83	1.00	2.25	1.00	3.00	2.00
3	3.00	2.00	7.00	2.32	2.34	4.39	1.98	1.92	4.81	2.99	1.00	1.74	3.00	3.90	1.00	2.72
4	1.00	3.03	2.32	7.00	4.98	3.00	2.31	2.40	2.14	2.31	3.20	1.06	2.00	1.23	4.00	2.05
5	2.00	5.01	2.34	4.98	7.00	3.00	3.32	0.42	2.15	3.33	1.18	1.08	2.00	1.24	4.00	1.06
6	2.57	3.00	4.39	3.00	3.00	7.00	2.41	1.32	4.58	3.41	2.00	0.66	3.00	3.49	2.00	1.67

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
7	2.00	3.00	1.98	2.31	3.32	2.41	7.00	0.91	1.83	3.99	1.00	1.75	2.00	2.92	1.00	1.74
8	1.00	0.00	1.92	2.40	0.42	1.32	0.91	7.00	2.74	1.91	4.00	3.66	2.00	2.83	2.00	2.65
9	2.15	2.00	4.81	2.14	2.15	4.58	1.83	2.74	7.00	3.82	1.00	2.92	4.00	2.91	2.00	3.91
10	2.00	3.00	2.99	2.31	3.33	3.41	3.99	1.91	3.82	7.00	1.00	2.75	1.00	1.91	3.00	2.74
11	0.00	1.83	1.00	3.20	1.18	2.00	1.00	4.00	1.00	1.00	7.00	2.00	2.08	1.00	2.00	1.00
12	0.22	1.00	1.74	1.06	1.08	0.66	1.75	3.66	2.92	2.75	2.00	7.00	2.00	0.83	3.00	3.99
13	1.00	2.25	3.00	2.00	2.00	3.00	2.00	2.00	4.00	1.00	2.08	2.00	7.00	2.00	0.00	1.00
14	2.06	1.00	3.90	1.23	1.24	3.49	2.92	2.83	2.91	1.91	1.00	0.83	2.00	7.00	0.00	0.82
15	1.00	3.00	1.00	4.00	4.00	2.00	1.00	2.00	2.00	3.00	2.00	3.00	0.00	0.00	7.00	3.00
…																3.40
120	2.24	2.00	2.72	2.05	1.06	1.67	1.74	2.65	3.91	2.74	1.00	3.99	1.00	0.82	3.00 3.40	7.00

Source: own study.

Negative results were then deleted as maximally differentiable. In the next stage, matrices of sums of similarity (non-differentiability) were calculated using the valued tolerance relation according to Formula 2 for all conditional attributes (Table 6).

Table 6

Indiscernibility sum matrix for conditional attributes

Sum	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	7.00	2.00	3.00	1.00	2.00	2.57	2.00	1.00	2.15	2.00	0.00	0.22	1.00	2.06	1.00
2	2.00	7.00	2.00	3.03	5.01	3.00	3.00	0.00	2.00	3.00	1.83	1.00	2.25	1.00	3.00
3	3.00	2.00	7.00	2.32	2.34	4.39	1.98	1.92	4.81	2.99	1.00	1.74	3.00	3.90	1.00
4	1.00	3.03	2.32	7.00	4.98	3.00	2.31	2.40	2.14	2.31	3.20	1.06	2.00	1.23	4.00
5	2.00	5.01	2.34	4.98	7.00	3.00	3.32	0.42	2.15	3.33	1.18	1.08	2.00	1.24	4.00
6	2.57	3.00	4.39	3.00	3.00	7.00	2.41	1.32	4.58	3.41	2.00	0.66	3.00	3.49	2.00
7	2.00	3.00	1.98	2.31	3.32	2.41	7.00	0.91	1.83	3.99	1.00	1.75	2.00	2.92	1.00
8	1.00	0.00	1.92	2.40	0.42	1.32	0.91	7.00	2.74	1.91	4.00	3.66	2.00	2.83	2.00
9	2.15	2.00	4.81	2.14	2.15	4.58	1.83	2.74	7.00	3.82	1.00	2.92	4.00	2.91	2.00
10	2.00	3.00	2.99	2.31	3.33	3.41	3.99	1.91	3.82	7.00	1.00	2.75	1.00	1.91	3.00
11	0.00	1.83	1.00	3.20	1.18	2.00	1.00	4.00	1.00	1.00	7.00	2.00	2.08	1.00	2.00
12	0.22	1.00	1.74	1.06	1.08	0.66	1.75	3.66	2.92	2.75	2.00	7.00	2.00	0.83	3.00
13	1.00	2.25	3.00	2.00	2.00	3.00	2.00	2.00	4.00	1.00	2.08	2.00	7.00	2.00	0.00
14	2.06	1.00	3.90	1.23	1.24	3.49	2.92	2.83	2.91	1.91	1.00	0.83	2.00	7.00	0.00
15	1.00	3.00	1.00	4.00	4.00	2.00	1.00	2.00	2.00	3.00	2.00	3.00	0.00	0.00	7.00
…															
120	2.24	2.00	2.72	2.05	1.06	1.67	1.74	2.65	3.91	2.74	1.00	3.99	1.00	0.82	3.00

SOURCE: own study

Module VI – setting decision-making rules within the abstraction classes for the given non-differentiability. This involves the classification of the rules of the analyzed transactions according to maximum similarity using the Boolean reasoning approach, i.e.: if…condition then … decision. As an example, in the presented simulation, the following formulas will be used for the first two transactions:

1. if (c_1 = 3) and (c_2 = 29,5) and (c_3 = 1) and (c_4 = 2) and (c_5 = 1) and (c_6 = 3) and (c_7 = 1) then (c= 4576),

2. if (c_1 = 2) and (c_2 = 75,2) and (c_3 = 4) and (c_4 = 3) and (c_5 = 1) and (c_6 = 7) and (c_7 = 1) then (c= 3125) … etc.

The next step was to determine the minimum level of similarity for the given set of data. Generally, a minimum similarity level of 80% was adopted due to the specifics of the real estate market, very

high differentiation of the manners of variable recording, and the quality and availability of information.

Module VII – in module VII, missing attributes are identified and edited on the basis of the quality of approximation. In this stage, the most approximate attribute is assigned to the object with missing data. This is determined on the basis of the results of the minimum similarity of the sum matrices. Hence, for the adopted simulation, it was agreed (Table 7) that for object 10, where the missing attribute was the location, real estate No. 101 will be the most approximate (similar). Therefore, the location is assigned at level 2, i.e. average. Below is an example of computations from the calculation sheet for real estate No. 10.

At this stage, objects with an approximation lower than the set maxima are reduced. In practice, we may deal with a case of duplication of approximating objects with maximum similarity; we then have to choose the object which has the greatest support in the number of objects belonging to the same decision-making rule. In another case, an object with minimum similarity may not be chosen at all; the object shall then be totally rejected or the minimum similarity threshold will be lowered.

In the adopted simulation, all missing attributes were determined, including the decisional attribute, i.e. the price. Table 8 presents the values of attributes determined for missing data.

Table 7

Assigning the missing attribute to real estate No. 10

	Real estate No. 10			
Trans. No.	Result for the matrix of sums for the cond. attribute after deleting the location	Remaining data above a similarity level of 80%	Result from the matrices of sums for cond. and dec. attributes after deleting the location	Remaining data above a similarity level of 80%
10	6.00	6.00	7.00	7.00
50	4.91	4.91	4.91	0.00
67	4.83	4.83	4.83	0.00
69	4.79	4.79	5.26	0.00
70	4.86	4.86	5.23	0.00
101	5.30	5.30	6.29	6.29

Source: own study.

Table 8

Approximate attribute values for property with incomplete data

Object with incomplete attributes	Approximated property	Incomplete data values							
		c_1	c_2	c_3	c_4	c_5	c_6	c_7	d
10	101	2							
15	100		42.8						
47	97			3					
82	45								2783
85	86				2				
116	115						7	0	

Source: own study.

Module VIII – calculation and analysis of the accuracy of approximation. In this module, the general concept of calculating the accuracy of approximation was used (Formula 5), bringing it down to the supplemented attribute and calculating the accuracy of supplemented data, accountingfor the number of known attributes and the value of the attribute of the approximate decision-making rule from the matrix of sums. In the presented simulation, the accuracy of approximation has been presented in Table 9 for all supplemented objects.

Table 9

Analysis of approximation quality

Object with incomplete attributes	Total number of attributes	Number of incomplete attributes	Number of known attributes	Value from the sum matrix of the approximate rule	Approximation quality
10	8	1	7	5.56	0.79
15	8	1	7	5.9	0.84
47	8	1	7	4.88	0.70
82	8	1	7	5.56	0.79
85	8	1	7	5.77	0.82
116	8	2	6	5.55	0.92

Source: own study.

Additionally, in the presented simulation, the effectiveness of the presented decision-making algorithm was checked by calculating the quality of supplemented data. To this end, the data previously deleted from the analyzed bases were compared with the data obtained in the course of the analysis (Table 10).

Table 10

Analysis of the efficiency of the decision-making algorithm

Object with incomplete attributes	Missing attribute (removed for the simulation)	Attribute determined on the basis of data filling procedure		Similarities quality	
10	Location	2	2	100%	
15	Surface area	44	42.8	97%	
47	Number of rooms	3	3	100%	
82	Price	2839.87	2782.67	98%	
85	Location on floor	2	2	100%	
	Age of Building	7	6	98%	99%
116	Cellar	0	0	100%	

SOURCE: own study

As can be seen from the table, 97% is the poorest result of similarity determination. Therefore, we can acknowledge that the developed model and the applied procedure bring satisfactory results and constitute an effective method of supplementing missing information.

Module IX – the last stage and finalization of analyses. Here missing data is supplied on the basis of the highest approximation, and we can then move on to the stage of decision-making, depending on the intended use of the database.

4. Summary

The authors prepared a simplified decision-making model of supplementing incomplete data in databases of real estate using the rough set theory. This theory enables one to create algorithms in a rather simple and clear manner, regardless of the volume of available information. It produces very positive results, both in small and large markets. As regards the analyses, the presented solution provides effective results; moreover, it may be flexibly modified and adjusted to a different analytical purpose, such as the valuation or specifying the intended use of real estate.

The aim of using these types of streamlined systems in real estate management is to increase the efficiency and scope of the decisions made.

5. References

BĄK J., NALEPKA A., 2006, *System wczesnego rozpoznania jako informacyjne wsparcie decyzji strategicznych (System of Early Recognition as Informational Support of Strategic Decisions)*, [in]: Informacja w zintegrowanej Europie (Information in Integrated Europe), ed. Borowiecki R. and Kwieciński M., Difin, Warsaw.

BRZEZICKA J., WIŚNIEWSKI R, 2014, *Wybrane postawy uczestników rynku wobec braków informacyjnych na rynku nieruchomości, (Selected Attitudes of Market Participants with Respect to Information Gaps in the Real Estate Market)*, Ekonomia XXI wieku, No. 2 (2), pp. 106-121.

BOLESTA-KUKUŁKA K., 2003, *Decyzje menedżerskie* (*Managerial decisions*), PWE, Warsaw.

HARRISON E. F., 1996, *A Process Perspective on Strategic Decision Making*, Management Decision, No. 34/1, pp. 46–53.

CELLMER R., SZCZEPANKOWSKA K., 2014, *Simulation Modeling in a Real Estate Market*, 9th International Conference "Environmental Engineering" Vilnius Gediminas Technical University, May 22-23, http://dx.doi.org/10.3846/enviro.2014.113.

CELLMER R., BELEJ M., ZROBEK S., ŠUBIC KOVAČ M., 2014, *Urban Land Value Maps – a Methodological Approach,* Geodetski Vestnik, Vol. 58, No. 3, pp. 535-55.

HATCH M. J., 2002, *Teoria organizacji* (*Theory of Organization*), PWN, Warsaw.

KONOWALCZUK J., RAMIAN T., 2013, *Potrzeby informacyjne inwestorów w procesie podejmowania decyzji inwestycyjnych na rynku nieruchomości –propozycja systematyki danych (Information Needs of Investors in the Investment Decision-Making Process on the Real Estate Market - Proposal of Systematizing Data)*. Studia Ekonomiczne (Economic Studies), Uniwersytet Ekonomiczny w Katowicach (Economic University in Katowice), No. 155, pp. 263-274.

PENC J., 2001, *Decyzje menedżerskie o sztuce zarządzania (Managerial Decisions on the Art of Management)*, Wydawnictwo C. H. Beck (C.H. Beck Publishing), Warsaw.

RADZEWICZ A., WALACIK M., 2014, *The Attributes of Sustainable Urban Development–Identification and Analysis on the Example of Olsztyn City*, 9th International Conference "Environmental Engineering" Vilnius Gediminas Technical University, May 22-23.

RENIGIER-BIŁOZOR M., 2010, *Supplementing Incomplete Databases on the Real Estate Market with the Use of the Rough Set Theory*. Acta Scientiarum Polonorum, Administratio Locorum, Vol. 9 (3) pp. 107-115.

RENIGIER-BIŁOZOR M., 2013, *Structure of a Decision Support Subsystem in Real Estate Management*. Folia Oeconomica Stetinensia, Vol. 13, Issue 1, 13 (21), pp. 56-75.

RENIGIER-BIŁOZOR M, WISNIEWSKI R., BIŁOZOR A., KAKLAUSKAS A., 2014, *Rating Methodology for Real Estate Markets – Poland Case Study*, International Journal of Strategic Property Management, Vol. 18, Issue 2, pp. 198-212.

WALACIK M., 2014, *Opracowanie zasad ustalania wysokości słusznego odszkodowania za nieruchomości przejęte na cele publiczne (Development of Principles for Establishing the Amount of Fair Compensation for Real Estate Expropriated for Public Purposes)* Wydawnictwo NDB (NDB Publishing), Olsztyn,